Name Disambiguation from link data in a collaboration graph

Baichuan Zhang, Tanay Kumar Saha, and Mohammad Al Hasan
Dept. of Computer and Info. Science, Indiana University - Purdue University Indianapolis
bz3@umail.iu.edu, tksaha@cs.iupui.edu, alhasan@cs.iupui.edu

Abstract—The entity disambiguation task partitions the records belonging to multiple persons with the objective that each decomposed partition is composed of records of a unique person. Existing solutions to this task use either biographical attributes, or auxiliary features that are collected from external sources, such as Wikipedia. However, for many scenarios, such auxiliary features are not available, or they are costly to obtain. Besides, the attempt of collecting biographical or external data sustains the risk of privacy violation. In this work, we propose a method for solving entity disambiguation task from link information obtained from a collaboration network. Our method is non-intrusive of privacy as it uses only the time-stamped graph topology of an anonymized network. Experimental results on two real-life academic collaboration networks show that the proposed method has satisfactory performance.